Ann. Inst. Statist. Math.
Vol. 46, No. 4, 757–768 (1994)

TUKEY'S LINEAR SENSITIVITY AND ORDER STATISTICS

H. N. NAGARAJA

Department of Statistics, The Ohio State University, 141 Cockins Hall, 1958 Neil Avenue, Columbus, OH 43210, U.S.A.

(Received April 2, 1993; revised January 14, 1994)

Abstract. Tukey (1965, *Proc. Nat. Acad. Sci. U.S.A.*, **53**, 127–134) introduced *linear sensitivity* as a measure of informativeness in a collection of order statistics. Here we study its general properties and discuss how it is related to the best linear unbiased estimator, Fisher information measure, and asymptotic relative efficiency. Also, we obtain explicit and asymptotic expressions for the linear sensitivity of a collection of consecutive order statistics from a location or from a scale family, and discuss its role in the comparison of L-estimators. We conclude our discussion with examples from uniform, exponential and normal populations.

Key words and phrases: Linear sensitivity, Fisher information measure, L-estimators, location family, scale family.

1. Introduction

Let Y be a random variable (rv) with cumulative distribution function (cdf) $F(y;\theta)$ where θ is a real valued parameter. Let $E(Y) = \mu(\theta)$ be a differentiable function of θ with derivative $\mu'(\theta)$, and $\mathrm{Var}(Y) = \sigma^2(\theta)$ be positive. Tukey (1965) defined *leverage* of Y as $\delta(Y;\theta) = \mu'(\theta)$, and *linear sensitivity* of Y as

$$\mathcal{S}(Y;\theta) = \frac{\{\delta(Y;\theta)\}^2}{\sigma^2(\theta)}, \tag{1.1}$$

which is taken to be zero when $\sigma^2(\theta)$ is infinite. He used $\mathcal{S}(Y;\theta)$ as a parametric measure of informativeness in Y about θ. We formally call it *Tukey's linear sensitivity* (TLS). Clearly the TLS is scale invariant and is one of the easiest measures to compute. Moreover, for two estimators W_1 and W_2 with the same mean, $\mathcal{S}(W_1;\theta)/\mathcal{S}(W_2;\theta)$ represents the efficiency of W_2 with respect to W_1.

Now suppose $\boldsymbol{Y}$ is a k-dimensional random vector with $\boldsymbol{Y} = (Y_1, \dots, Y_k)'$ having mean vector $E(\boldsymbol{Y}) = \boldsymbol{\mu} = (\mu_1(\theta), \dots, \mu_k(\theta))'$ and covariance matrix $C(\boldsymbol{Y}) = \boldsymbol{\Sigma}$ whose elements σ_{ij} possibly depend on θ. Let $\delta_i(\theta) = \mu_i'(\theta)$, $1 \le i \le k$ and $\boldsymbol{c} = (c_1, \dots, c_k)'$ where the c_i's may depend on θ. Then Tukey defines leverage of $\boldsymbol{c}'\boldsymbol{Y}$ as $\delta(\boldsymbol{c}'\boldsymbol{Y};\theta) = \sum_{i=1}^k c_i\delta_i(\theta) = \boldsymbol{c}'\boldsymbol{\Delta}$ where $\boldsymbol{\Delta} = (\delta_1, \dots, \delta_k)'$. With this convention, the TLS of the vector $\boldsymbol{Y}$ is defined to be $\mathcal{S}(\boldsymbol{Y};\theta) = \sup_c\{\mathcal{S}(\boldsymbol{c}'\boldsymbol{Y};\theta)\}$. In other words,

$$\mathcal{S}(\boldsymbol{Y};\theta) = \sup_{\boldsymbol{c}}\{(\boldsymbol{c}'\boldsymbol{\Delta})^2/\boldsymbol{c}'\boldsymbol{\Sigma}\boldsymbol{c}\}. \tag{1.2}$$

Tukey studied some of the properties of $\mathcal{S}(\boldsymbol{Y};\theta)$ when $\boldsymbol{Y} = (Y_{r:n},\dots,Y_{k:n})'$, $1 \le r < k \le n$, where $Y_{i:n}$ is the i-th order statistic from a random sample of size n from the absolutely continuous cdf $F(y;\theta)$. In that context $\mathcal{S}$ provides a meaningful measure of efficiency of unbiased L-estimators of location and scale parameters. Possibly because of its unusual place of appearance, Tukey's work went unnoticed even though the literature on linear estimation based on order statistics is quite huge (Balakrishnan and Cohen (1991), Chapter 4).

In this paper, we explore general properties of the measure $\mathcal{S}$. We obtain below an explicit expression for $\mathcal{S}(\boldsymbol{Y};\theta)$ as a function of $\boldsymbol{\Delta}$ and $\boldsymbol{\Sigma}$. In Section 2 we explore some desirable properties of $\mathcal{S}(\boldsymbol{Y};\theta)$ and discuss how it is related to best linear unbiased estimator (BLUE) of θ, to the Fisher information measure (FIM) in $\boldsymbol{Y}$, $\mathcal{I}(\boldsymbol{Y};\theta)$, and to the concept of asymptotic relative efficiency (ARE). In Section 3 we observe some deficiencies of $\mathcal{S}$. In Section 4, we obtain expressions for and approximations to $\mathcal{S}(\boldsymbol{Y};\theta)$, when $\boldsymbol{Y}$ is a vector of consecutive order statistics from a random sample, and θ is either a location or a scale parameter. In both these cases, $\mathrm{Var}(\hat{\theta}) = \{\mathcal{S}(\boldsymbol{Y};\theta)\}^{-1}$, where $\hat{\theta}$ is the BLUE of θ. Thus, for any L-statistic T that is unbiased for θ, $\mathcal{S}(T;\theta)/\mathcal{S}(\boldsymbol{Y};\theta)$ represents the efficiency of T with respect to the BLUE. The last section is devoted to examples on TLS of order statistics. There we compute $\mathcal{S}(\boldsymbol{Y};\theta)$ for uniform, exponential and normal populations and compare it with $\mathcal{I}(\boldsymbol{Y};\theta)$ whenever the latter is defined.

The following lemma comes in handy in our exploration of the properties of TLS.

LEMMA 1.1. *When $\boldsymbol{\Sigma}$ is positive definite, $\mathcal{S}(\boldsymbol{Y};\theta) = \boldsymbol{\Delta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}$ and the linear function of $\boldsymbol{Y}$ with maximum TLS is of the form $d(\theta)\boldsymbol{\Delta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}$, where $d(\theta)$ is a scalar function of θ.*

PROOF. From a well-known result which is essentially a version of the Cauchy-Schwarz inequality (Rao (1973), p. 60), it follows that $\sup_c\{((\boldsymbol{c}'\boldsymbol{\Delta})^2/\boldsymbol{c}'\boldsymbol{\Sigma}\boldsymbol{c})\} = \boldsymbol{\Delta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}$. That result also identifies the vector $\boldsymbol{c}$ corresponding to the maximum value as being proportional to $\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}$. □

When $\boldsymbol{\Sigma}$ is just positive semidefinite, one can find a $\boldsymbol{c}$ for which $\boldsymbol{c}'\boldsymbol{\Sigma}\boldsymbol{c}$ is zero, and hence $\mathcal{S}(\boldsymbol{Y};\theta)$ will be undefined.

2. Some connections and desirable features of $\mathcal{S}$

2.1 *Relationship with the BLUE*

Suppose $d(\theta)$ can be chosen such that $\boldsymbol{a}' = d(\theta)\boldsymbol{\Delta}'\boldsymbol{\Sigma}^{-1}$ is free of θ. Then, clearly $\boldsymbol{a}'\boldsymbol{Y}$ is the BLUE of $\boldsymbol{a}'\boldsymbol{\mu}$ since for any other linear unbiased estimator $\boldsymbol{c}'\boldsymbol{Y}$, $\boldsymbol{c}'\boldsymbol{\Delta} = \boldsymbol{a}'\boldsymbol{\Delta}$ and $\{(\boldsymbol{c}'\boldsymbol{\Delta})^2/\boldsymbol{c}'\boldsymbol{\Sigma}\boldsymbol{c}\} \le \{(\boldsymbol{a}'\boldsymbol{\Delta})^2/\boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a}\}$. Further, $\mathcal{S}(\boldsymbol{a}'\boldsymbol{Y};\theta) = \mathcal{S}(\boldsymbol{Y};\theta)$.

In the usual linear model setting, let $\boldsymbol{Y} = \boldsymbol{A\theta} + \boldsymbol{e}$, where $\boldsymbol{e}$ is a random vector with zero mean vector and $C(\boldsymbol{e}) = \boldsymbol{\Sigma}$ is positive definite. Let $\hat{\theta}$ be the BLUE of θ, where θ is a linear function of components of $\boldsymbol{\theta}$. Then for T, an unbiased linear estimator of θ, $\{\mathcal{S}(\hat{\theta};\theta)/\mathcal{S}(T;\theta)\} \leq 1$ and the ratio represents the efficiency of T with respect to $\hat{\theta}$.

We will explore the special role of TLS in the linear estimation of location and scale parameters based on order statistics in Section 4.

2.2 *Relationship with the Fisher information measure (FIM)*

Let $\mathcal{I}(T;\theta)$ denote the FIM contained in a statistic T and $\mu(\theta)$ be its mean. Whenever $\mathcal{I}(T;\theta)$ is positive and $\mu'(\theta)$ exists, from Cramér-Rao inequality it is known that $\mathrm{Var}(T) \geq \{(\mu'(\theta))^2/\mathcal{I}(T;\theta)\}$. Thus, as Tukey ((1965), p. 128) notes, if $\mathcal{I}(T;\theta)$ exists, it provides an upper bound for $\mathcal{S}(T;\theta)$. The two measures coincide if, and only if (iff) equality is attained in the Cramér-Rao inequality. From Wijsman (1973) (see also Lehmann (1983), p. 123) it then follows that $\mathcal{I}(T;\theta) = \mathcal{S}(T;\theta)$ iff T has a one-parameter exponential family density.

We have seen above that if there exists $d(\theta)$ such that $\boldsymbol{a}' = d(\theta)\boldsymbol{\Delta}'\boldsymbol{\Sigma}^{-1}$ is free of θ, $\mathcal{S}(\boldsymbol{Y};\theta) = \mathcal{S}(T;\theta)$ where T is the BLUE of $\boldsymbol{a}'\boldsymbol{\mu}$. Since $\mathcal{I}(\boldsymbol{Y};\theta) \geq \mathcal{I}(T;\theta)$ for any statistic T, we can then conclude that $\mathcal{I}(\boldsymbol{Y};\theta)$ provides an upper bound for $\mathcal{S}(\boldsymbol{Y};\theta)$. The two coincide iff $\boldsymbol{a}'$ is free of θ and T belongs to a one-parameter exponential family. This is the case, for example, when $Y_1, \ldots, Y_n$ themselves form a random sample from a one-parameter exponential family density $f(y;\theta) = h(y)\exp\{\eta(\theta)y - B(\theta)\}$. It is easy to see that, in this case $\{\eta'(\theta)\}^{-1}\boldsymbol{\Delta}'\boldsymbol{\Sigma}^{-1} = (1, \ldots, 1)$ can be chosen as $\boldsymbol{a}'$ and $T = \sum_{i=1}^{n} Y_i$.

2.3 *Connection with ARE*

Suppose T_n is a statistic based on a random sample of size n, which is asymptotically normal with mean $\mu_n(\theta)$ and variance $\sigma_n^2(\theta)$. In the inference literature, the *efficacy* of T_n is defined to be $e_n(\theta) = \{(\mu_n'(\theta))^2/\sigma_n^2(\theta)\}$, assuming $\mu_n(\theta)$ is differentiable (Pratt and Gibbons (1981), p. 378). From (1.1), it is clear that $e_n(\theta)$ *is* $\mathcal{S}(T_n;\theta)$. The asymptotic efficacy of T_n is $e(T) = \lim_{n\to\infty}(e_n(\theta)/n)$. The ARE of T_n relative to another sequence T_n^* is defined to be $\mathrm{ARE} = e(T)/e(T^*)$. This means $\mathrm{ARE} = \lim_{n\to\infty}\{\mathcal{S}(T_n;\theta)/\mathcal{S}(T_n^*;\theta)\}$.

2.4 *Desirable properties*

The TLS has some of the desirable properties expected of a measure of information. It is nonnegative whenever it is defined, and for rv's X and Y, $\mathcal{S}(X, Y;\theta) = \mathcal{S}(X;\theta) + \mathcal{S}(Y;\theta)$ whenever X and Y are independent, or are just uncorrelated. The latter property is known as weak additivity.

THEOREM 2.1. (Convexity) *If* $F_3(\boldsymbol{y};\theta) = \alpha F_1(\boldsymbol{y};\theta) + (1-\alpha)F_2(\boldsymbol{y};\theta)$, $0 < \alpha < 1$,

$$\mathcal{S}_3(\boldsymbol{Y};\theta) \leq \alpha\mathcal{S}_1(\boldsymbol{Y};\theta) + (1-\alpha)\mathcal{S}_2(\boldsymbol{Y};\theta), \tag{2.1}$$

where $\mathcal{S}_j(\boldsymbol{Y};\theta)$ *is the TLS of* $\boldsymbol{Y}$ *when its cdf is* $F_j(\boldsymbol{y};\theta)$, $j = 1, 2, 3$. *Thus,* $\mathcal{S}$ *is convex.*

PROOF. Let $\boldsymbol{\mu}_j$ denote the mean vector, $\boldsymbol{\Delta}_j$, its derivative (with respect to θ), and $\boldsymbol{\Sigma}_j$ be the covariance matrix of $\boldsymbol{Y}$ when $F = F_j$, $j = 1, 2, 3$. Now, for an arbitrary vector $\boldsymbol{c}$, let $\boldsymbol{c}'\boldsymbol{\Delta}_j = \lambda_j$ and $\boldsymbol{c}'\boldsymbol{\Sigma}_j\boldsymbol{c} = \gamma_j$. We will show that

$$\{\lambda_3^2/\gamma_3\} \leq (\alpha/\gamma_1)\lambda_1^2 + \{(1-\alpha)/\gamma_2\}\lambda_2^2. \tag{2.2}$$

Since $\boldsymbol{\Delta}_3 = \alpha\boldsymbol{\Delta}_1 + (1-\alpha)\boldsymbol{\Delta}_2$, $\lambda_3 = \alpha\lambda_1 + (1-\alpha)\lambda_2$. Further, on using the fact that $\boldsymbol{\Sigma}_3 = \alpha\boldsymbol{\Sigma}_1 + (1-\alpha)\boldsymbol{\Sigma}_2 + \alpha(1-\alpha)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'$, it follows that $\gamma_3 = \alpha\gamma_1 + (1-\alpha)\gamma_2 + \alpha(1-\alpha)\{\boldsymbol{c}'(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{c}\}$. Using these representations, it can be seen that (2.2) holds iff, $\{\boldsymbol{c}'(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{c}\}[(\alpha/\gamma_1)\lambda_1^2 + \{(1-\alpha)/\gamma_2\}\lambda_2^2] + \{(\gamma_2/\gamma_1)^{1/2}\lambda_1 - (\gamma_1/\gamma_2)^{1/2}\lambda_2\}^2 \geq 0$. This is obviously true. In view of (1.2), we then can conclude that $\mathcal{S}$ is convex.

Equality is achieved in (2.2) if $\lambda_1 = \lambda_2 = 0$, or if $\lambda_1 = \lambda_2$ and $\gamma_1 = \gamma_2$. As a consequence, we can conclude that equality holds in (2.1) if either the mean vectors are free of θ or when F_1 and F_2 have common mean vectors and covariance matrices. □

Let $\boldsymbol{Y}' = (\boldsymbol{Y}_1', \boldsymbol{Y}_2')$. From the definition, it is evident that $\mathcal{S}(\boldsymbol{Y};\theta) \geq \mathcal{S}(\boldsymbol{Y}_1;\theta)$ implying the monotonicity of $\mathcal{S}$. What about the increase in the amount of TLS? This question is answered by the following result.

THEOREM 2.2. (Monotonicity) *Let* $\boldsymbol{\Delta}' = (\boldsymbol{\Delta}_1', \boldsymbol{\Delta}_2')$ *and* $\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$ *where the partitions correspond to the partitioning of* $\boldsymbol{Y}$ *into* $\boldsymbol{Y}_1$ *and* $\boldsymbol{Y}_2$. *Suppose* $\boldsymbol{\Sigma}$ *is positive definite and let* $\boldsymbol{\Sigma}_{2.1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$. *Then,*

$$\mathcal{S}(\boldsymbol{Y};\theta) = \mathcal{S}(\boldsymbol{Y}_1;\theta) + (\boldsymbol{\Delta}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Delta}_1)'\boldsymbol{\Sigma}_{2.1}^{-1}(\boldsymbol{\Delta}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Delta}_1). \tag{2.3}$$

PROOF. This result follows from Lemma 1.1 and the discussion about the joint and conditional distributions associated with a multivariate normal random vector. (See, for example, Anderson (1984), p. 35.) □

Theorem 2.2 provides an explicit expression for the increase in TLS when additional observations are added to the data set. From the relation (2.3) it is evident that the increase can be zero only when $\boldsymbol{\Delta}_2 = \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Delta}_1$. If $\boldsymbol{\Sigma}$ is free of θ, this is possible iff $E(\boldsymbol{Y}_2) - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}E(\boldsymbol{Y}_1)$ does not involve θ. This can happen when $\boldsymbol{Y}$ is multivariate normal, and in the context of order statistics also such a situation can arise. See Examples 1 and 2 in Section 5.

3. Deficiencies of $\mathcal{S}$

Ferentinos and Papaioannou (1981) provide an extensive list of desirable properties expected of a measure of information. We now examine TLS and note that it has some deficiencies in the sense it fails to satisfy some of these properties.

For a rv X, let $\mathcal{J}(X;\theta)$ be a measure of information contained in X about θ. It has *conditional inequality* property if for any rv Y, $\mathcal{J}(Y;\theta) \geq \mathcal{J}(Y \mid X;\theta)$. This property does not hold for $\mathcal{S}$. For example, let X be uniformly distributed over

$(0,2)$ and given $X = x$, let Y be normal with mean θ and variance x, to be written $N(\theta, x)$. Then $E(Y) = E(Y \mid X) = \theta$, $\mathrm{Var}(Y \mid X) = X$, and $\mathrm{Var}(Y) = E(\mathrm{Var}(Y \mid X)) + \mathrm{Var}(E(Y \mid X)) = 1$. Hence $\mathcal{S}(Y;\theta) = 1$ whereas $\mathcal{S}(Y \mid X;\theta) = 1/X$ which can clearly exceed $\mathcal{S}(Y;\theta)$.

The information measure $\mathcal{J}$ is *strongly additive* if $\mathcal{J}(X,Y;\theta) = \mathcal{J}(X;\theta) + \mathcal{J}(Y \mid X;\theta)$, and is *subadditive* if $\mathcal{J}(X,Y;\theta) \le \mathcal{J}(X;\theta) + \mathcal{J}(Y;\theta)$. These properties are not satisfied by $\mathcal{S}$. In the above example, $\mathcal{S}(X,Y;\theta) = \mathcal{S}(Y;\theta)$ while $\mathcal{S}(X;\theta) + \mathcal{S}(Y \mid X;\theta) = \mathcal{S}(Y \mid X;\theta)$. Hence the strong additivity property does not hold.

To substantiate the other claim, take $\boldsymbol{Y} = (Y_1, Y_2)'$ and let $\mathcal{S}_i = \mathcal{S}(Y_i;\theta)$, $i = 1, 2$. Let ρ be the correlation between Y_1 and Y_2. Then,

$$\begin{aligned} \mathcal{S}(Y_1, Y_2;\theta) &= \boldsymbol{\Delta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta} \\ &= (1-\rho^2)^{-1}\{\mathcal{S}_1 + \mathcal{S}_2 - 2\rho\sqrt{\mathcal{S}_1\mathcal{S}_2}\}. \end{aligned} \tag{3.1}$$

This can exceed $\mathcal{S}_1 + \mathcal{S}_2$, for example when $E(Y_2)$ is free of θ, violating subadditivity property.

Now, suppose $\mathcal{S}_1 = \mathcal{S}_2 = \mathcal{S}$, say. Then, from (3.1), it follows that $\mathcal{S}(Y_1, Y_2;\theta) = 2\mathcal{S}(1+\rho)^{-1}$, while $\mathcal{S}(Y_1;\theta) + \mathcal{S}(Y_2;\theta) = 2\mathcal{S}$. Thus, $\mathcal{S}(Y_1, Y_2;\theta) \ge (<)\ \mathcal{S}(Y_1;\theta) + \mathcal{S}(Y_2;\theta)$, according as $\rho \le (>)\ 0$. As to be expected, equality is achieved here if $\rho = 0$.

Another desirable feature is the *maximal information* property which holds for $\mathcal{J}$ if $\mathcal{J}(\boldsymbol{Y};\theta) \ge \mathcal{J}(T;\theta)$ for any statistic T based on $\boldsymbol{Y}$. This is satisfied by $\mathcal{S}$ in a restricted sense. If T is a linear function of the components of $\boldsymbol{Y}$, then clearly $\mathcal{S}(T;\theta) \le \mathcal{S}(\boldsymbol{Y};\theta)$. However, when T is a nonlinear function, this inequality could be reversed. For example, when Y is $N(0,\theta)$, $\mathcal{S}(Y;\theta) = 0$ whereas $\mathcal{S}(Y^2;\theta) = (2\theta^2)^{-1}$ is always positive.

The information measure $\mathcal{J}$ is invariant under sufficient transformations if $\mathcal{J}(\boldsymbol{Y};\theta) = \mathcal{J}(T(\boldsymbol{Y});\theta)$ iff $T(\boldsymbol{Y})$ is a sufficient statistic. The TLS fails to satisfy this except when T is a linear function of components of $\boldsymbol{Y}$. In general $\mathcal{S}(T;\theta)$ could exceed or be less than $\mathcal{S}(\boldsymbol{Y};\theta)$. An extreme case where $\mathcal{S}(T;\theta) = 0$ and $\mathcal{S}(\boldsymbol{Y};\theta) > 0$ is also possible. Tukey ((1965), p. 128) provides such an example. Thus, based on previous results we conclude that TLS does not satisfy the main properties of measures of information. Therefore TLS is not a good measure of information. However, while dealing with a single random variable, it provides a lower bound on FIM.

4. Linear sensitivity of order statistics

Tukey (1965) introduced the concept of linear sensitivity to describe the amount of information contained in single or a block of consecutive order statistics. He described the information content of an order statistic $Y_{i:n}$ in three ways: (a) $\mathcal{S}(Y_{i:n};\theta)$, the linear sensitivity of $Y_{i:n}$, (b) $\mathcal{S}_R(Y_{i:n};\theta)$, the increase in linear sensitivity when $Y_{i:n}$ is added on the right, given by $\mathcal{S}(\boldsymbol{Y_1};\theta) - \mathcal{S}(\boldsymbol{Y_1^*};\theta)$, where $\boldsymbol{Y_1} = (Y_{1:n}, \ldots, Y_{i:n})'$ and $\boldsymbol{Y_1^*} = (Y_{1:n}, \ldots, Y_{i-1:n})'$, (c) $\mathcal{S}_L(Y_{i:n};\theta)$, the increase in linear sensitivity when $Y_{i:n}$ is added on the left, given by $\mathcal{S}(\boldsymbol{Y_2};\theta) - \mathcal{S}(\boldsymbol{Y_2^*};\theta)$, where $\boldsymbol{Y_2} = (Y_{i:n}, \ldots, Y_{n:n})'$ and $\boldsymbol{Y_2^*} = (Y_{i+1:n}, \ldots, Y_{n:n})'$. A block of order statistics $\boldsymbol{Y} = (Y_{r:n}, \ldots, Y_{k:n})'$ can play the role of $Y_{i:n}$ in each of these situations.

Suppose the parent distribution is absolutely continuous with probability density function (pdf) $f(y;\theta)$ where θ is either a location or a scale parameter. We now obtain expressions for $\mathcal{S}(\boldsymbol{Y};\theta)$ when $\boldsymbol{Y}$ is a vector of selected order statistics, and note that, in both the cases,

$$\text{Var}(\hat{\theta}) = \{\mathcal{S}(\boldsymbol{Y};\theta)\}^{-1}, \tag{4.1}$$

where $\hat{\theta}$ is the BLUE of θ based on $\boldsymbol{Y}$. In contrast, $\mathcal{I}(\boldsymbol{Y};\theta)$ is rather complicated to compute, even when it exists. When the components of $\boldsymbol{Y}$ form a block of consecutive order statistics, we also obtain asymptotic expressions for $\mathcal{S}(\boldsymbol{Y};\theta)$.

4.1 *Location family distributions*

Let $f(y;\theta) = g(y-\theta)$ and $Z_{i:n} = Y_{i:n} - \theta$ represent the order statistic from the standardized distribution with pdf g and cdf G. Then $\boldsymbol{Y}$ has the same distribution as $\boldsymbol{Z} + \theta\mathbf{1}$, where $\boldsymbol{Z}$ is the vector of standardized order statistics and $\mathbf{1}$ is the vector of 1's. Let $E(\boldsymbol{Z}) = \boldsymbol{\alpha}$ and $C(\boldsymbol{Z}) = \boldsymbol{\Sigma}$. Consequently, $E(\boldsymbol{Y}) = \boldsymbol{\alpha} + \theta\mathbf{1}$ and $C(\boldsymbol{Y}) = \boldsymbol{\Sigma}$. Hence, from Lemma 1.1, we have

$$\mathcal{S}(\boldsymbol{Y};\theta) = \mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1}. \tag{4.2}$$

Lloyd (1952) has shown that the BLUE of θ is $\hat{\theta} = \mathbf{1}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}/\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1}$ and $\text{Var}(\hat{\theta}) = (\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1})^{-1}$. Consequently, (4.1) holds.

For n large, using the theory developed for optimal asymptotic estimation by order statistics, we can find an approximation to $\mathcal{S}(\boldsymbol{Y};\theta)$ when $\boldsymbol{Y} = (Y_{r:n}, \ldots, Y_{k:n})'$, a vector of consecutive order statistics. Now assume $(r/n) \to p_1$, $(k/n) \to p_2$, $0 < p_1 < p_2 < 1$. From (4.2), and (9.7.3a) of David ((1981), p. 278), it follows that

$$\mathcal{S}(\boldsymbol{Y};\theta) \cong n\left\{\int_{G^{-1}(p_1)}^{G^{-1}(p_2)} \left\{\frac{g'(z)}{g(z)}\right\}^2 dG(z) + \frac{g^2(G^{-1}(p_1))}{p_1} + \frac{g^2(G^{-1}(p_2))}{1-p_2}\right\}, \tag{4.3}$$

where $\cong$ stands for 'approximately equal'. Tukey ((1965), p. 132) has given expressions comparable to (4.3) using a heuristic argument.

If the pdf g is smooth around the extreme points of its support so that $\{g^2(G^{-1}(p))/p\}$ approaches 0 as p approaches 0 or 1, we can obtain an approximation to $\mathcal{S}(\boldsymbol{Y};\theta)$ using (4.3) when $\boldsymbol{Y}$ represents a singly Type II censored sample. For example, for the right censored sample with $(k/n) \to p_2$, the middle term on the right side of (4.3) vanishes. In the uncensored case, one has $\mathcal{S}(\boldsymbol{Y};\theta) \cong n\int_{-\infty}^{\infty}\{g'(z)/g(z)\}^2 dG(z) = \mathcal{I}(\boldsymbol{Y};\theta)$, the FIM contained in the entire sample. In the case of a single order statistic $Y_{k:n}$, if $(k/n) \to p$, $0 < p < 1$, $\mathcal{S}(Y_{k:n};\theta) \cong ng^2(G^{-1}(p))\{p(1-p)\}^{-1}$.

4.2 *Scale family distributions*

The discussion here runs parallel to the location family case. With $f(y;\theta) = \theta^{-1}g(y/\theta)$, $\theta > 0$, let $Z_{i:n} = Y_{i:n}/\theta$ represent the order statistics from the standardized distribution. We then have $E(\boldsymbol{Y}) = \theta\boldsymbol{\alpha}$ and $C(\boldsymbol{Y}) = \boldsymbol{\Sigma} = \theta^2\boldsymbol{\Sigma}_0$ where $\boldsymbol{\Sigma}_0 = C(\boldsymbol{Z})$. Hence, we obtain

$$\mathcal{S}(\boldsymbol{Y};\theta) = \boldsymbol{\alpha}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\alpha} = (\boldsymbol{\alpha}'\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\alpha})/\theta^2. \tag{4.4}$$

The BLUE of θ is $\hat{\theta} = \boldsymbol{\alpha}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}/\boldsymbol{\alpha}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\alpha}$ and once again (4.1) holds.

As before, an approximation to $\mathcal{S}(\boldsymbol{Y};\theta)$ in (4.4) can be obtained when $\boldsymbol{Y} = (Y_{r:n},\dots,Y_{k:n})'$, a vector of consecutive order statistics. From (9.7.3c) of David ((1981), p. 278), we obtain as $(r/n) \to p_1$, $(k/n) \to p_2$, $0 < p_1 < p_2 < 1$,

$$\lim_{n\to\infty}\frac{1}{n}\mathcal{S}(\boldsymbol{Y};\theta) = \int_{p_1}^{p_2}\left\{1 + G^{-1}(u)\frac{g'(G^{-1}(u))}{g(G^{-1}(u))}\right\}^2 du + \frac{\{g(G^{-1}(p_1))G^{-1}(p_1)\}^2}{p_1} + \frac{\{g(G^{-1}(p_2))G^{-1}(p_2)\}^2}{1-p_2}. \tag{4.5}$$

If g has smooth tails so that the last two terms above vanish as $p_1 \to 0$ and $p_2 \to 1$, we can conclude that, for the whole sample,

$$\mathcal{S}(\boldsymbol{Y};\theta) \cong n\int_{-\infty}^{\infty}\left\{1 + z\frac{g'(z)}{g(z)}\right\}^2 dG(z),$$

which once again represents $\mathcal{I}(\boldsymbol{Y};\theta)$.

4.3 *Remarks*

(1) Using the expressions and approximations for $\mathcal{S}(\boldsymbol{Y};\theta)$, one can obtain similar results for $\mathcal{S}_R(\boldsymbol{Y};\theta)$ and $\mathcal{S}_L(\boldsymbol{Y};\theta)$.

(2) From (4.2) and (4.4) it is clear that to compute $\mathcal{S}(\boldsymbol{Y};\theta)$ we need the elements of $\boldsymbol{\Sigma}^{-1}$. When σ_{ij} is of the form a_ib_j for $i \le j$, all the entries of $\boldsymbol{\Sigma}^{-1}$ are zeros except for the ones on the main diagonal and the ones on the two adjacent off-diagonals. Further the nonzero entries are simple functions of the a_i's and b_j's (see Arnold *et al.* (1992), pp. 174–175, for an explicit form). This is the case with the families of distributions consisting of power function, Pareto and exponential distributions. Further, the (i,j)-th element of the asymptotic covariance matrix of selected order statistics is of the form a_ib_j (see, for example, Arnold *et al.* (1992), p. 226). Thus, it is easy to compute $\mathcal{S}(\boldsymbol{Y};\theta)$ for some distributions with a location or scale parameter, and it is easily approximated for all distributions from these families.

(3) Mosteller (1946) showed that the joint asymptotic distribution of k selected central order statistics is multivariate normal under some mild conditions on the cdf F. For the location and scale families, Ogawa (1951) used Mosteller's work to determine the optimal choices for order statistics so that one maximizes

$\mathcal{I}(\boldsymbol{Y};\theta)$. From his results (summarized in Ogawa (1962)) it is evident that under the assumption of asymptotic multivariate normality, maximizing $\mathcal{I}(\boldsymbol{Y};\theta)$ is equivalent to minimizing $\mathrm{Var}(\hat{\theta})$ where $\hat{\theta}$ is the BLUE of θ. It then follows from (4.1) that, for both the location and scale family distributions, the choice would correspond to the maximum value of $\mathcal{S}(\boldsymbol{Y};\theta)$.

(4) For large samples, the right sides of (4.3) and (4.5) also provide close approximations to the FIM in $\boldsymbol{Y}$. Recently, for general one-parameter family, Takahashi and Sugiura (1989) have investigated the rate of convergence of the sample FIM in Type II censored samples to the limiting Fisher information.

5. Examples

Throughout this section, $\boldsymbol{Y}$ represents the vector of consecutive order statistics given by $(Y_{r:n},\ldots,Y_{k:n})'$.

Example 1. (The Uniform Distribution) In the location family setting, let $f(y;\theta) = 1$, $\theta - 0.5 \le y \le \theta + 0.5$. Here $\alpha_i = i/(n+1)$, and $\sigma_{ij} = i(n-j+1)/\{(n+1)^2(n+2)\}$, $i \le j$, is of the form $a_i b_j$. On using the expression for $\boldsymbol{\Sigma}^{-1}$ given in David ((1981), p. 172), we conclude that $\mathcal{S}(\boldsymbol{Y};\theta) = \mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1} = (n+1)(n+2)\{r^{-1} + (n-k+1)^{-1}\} = \mathcal{S}(Y_{r:n}, Y_{k:n};\theta)$.

When θ is the scale parameter, take $f(y;\theta) = 1/\theta$, $0 \le y \le \theta$. In this case, $\alpha_i = i/(n+1)$ as before, and $\sigma_{ij} = \theta^2 i(n-j+1)/\{(n+1)^2(n+2)\}$, $i \le j$. From (4.4), on simplification, we obtain $\mathcal{S}(\boldsymbol{Y};\theta) = k(n+2)/\{(n-k+1)\theta^2\}$. In this case $\mathcal{S}(\boldsymbol{Y};\theta) = \mathcal{S}(Y_{k:n};\theta)$. Hence adding observations on the left does not increase the information about θ. In other words, $\mathcal{S}_L(Y_{i:n};\theta) = 0$, justifying the remarks made at the end of Section 2. The BLUE of θ is $\hat{\theta} = (n+1)Y_{k:n}/k$.

Note that $\mathcal{I}(\boldsymbol{Y};\theta)$ does not exist in both these settings.

Example 2. (The Exponential Distribution) (a) In the location family setting we let $f(y;\theta) = \exp(-(y-\theta))$, $y \ge \theta$. Let us define

$$\alpha_{i:n} = \sum_{j=1}^{i}(n-j+1)^{-1} \quad \text{and} \quad \beta_{i:n} = \sum_{j=1}^{i}(n-j+1)^{-2}. \tag{5.1}$$

Since $\sigma_{ij} = \sigma_{ii} = \beta_{i:n}$ for $i \le j$, and thus is of the form $a_i b_j$, $\boldsymbol{\Sigma}^{-1}$ can easily be computed. On simplification we obtain $\mathcal{S}(\boldsymbol{Y};\theta) = (\beta_{r:n})^{-1} = \mathcal{S}(Y_{r:n};\theta)$, the TLS of the sufficient statistic. This also means $\mathcal{S}_R(Y_{i:n};\theta) = 0$.

(b) In the scale family setting, $f(y;\theta) = \theta^{-1}\exp(-(y/\theta))$, $y \ge 0$. In this case, $\alpha_{i:n}$'s are the same as in (5.1), but $\sigma_{ij} = \theta^2\beta_{i:n}$ for $i \le j$. On using (4.4) we obtain

$$\begin{aligned} \mathcal{S}(\boldsymbol{Y};\theta) &= \frac{1}{\theta^2}\left\{\frac{\alpha_{r:n}^2}{\beta_{r:n}} + (k-r)\right\} \\ &= \mathcal{S}(Y_{r:n};\theta) + (k-r)\theta^{-2}. \end{aligned} \tag{5.2}$$

This means $\mathcal{S}_R(\boldsymbol{Y};\theta)$ increases linearly, whereas the growth of $\mathcal{S}_L(\boldsymbol{Y};\theta)$ is nonlinear. Further, the linear function of the components of $\boldsymbol{Y}$ which contains all the information in $\boldsymbol{Y}$ is $(\alpha_{r:n}/\beta_{r:n})Y_{r:n} + \sum_{j=r+1}^{k}(n-j+1)(Y_{j:n} - Y_{j-1:n})$.

One can also compute $\mathcal{I}(\boldsymbol{Y};\theta)$. In fact, it is easily seen that

$$\mathcal{I}(\boldsymbol{Y};\theta) = \mathcal{I}(Y_{r:n};\theta) + (k-r)\theta^{-2}, \tag{5.3}$$

where $\mathcal{I}(Y_{r:n};\theta)$ is rather messy to compute. From Arnold *et al.* ((1992), p. 166) it follows that

$$\theta^2\mathcal{I}(Y_{r:n};\theta) = \begin{cases} 1 & \text{for } r=1 \\ 1+2n(n-1)\sum_{j=0}^{\infty}(n+j)^{-3} & \text{for } r=2 \\ 1+n(n-r+1)(r-2)^{-1}\{\alpha^2_{r-2:n-1}+\beta_{r-2:n-1}\} & \\ & \text{for } r\geq 3, \end{cases} \tag{5.4}$$

where the α's and β's are given in (5.1). From (5.2)–(5.4), it is clear that $\mathcal{S}(\boldsymbol{Y};\theta)$ and $\mathcal{I}(\boldsymbol{Y};\theta)$ coincide when $r=1$, as they should be, since $Y_{1:n}$ belongs to the one-parameter exponential family; otherwise, the difference is nothing but $\mathcal{S}(Y_{r:n};\theta)-\mathcal{I}(Y_{r:n};\theta)$.

Now let us see how close $\mathcal{S}(Y_{r:n};\theta)$ and $\mathcal{I}(Y_{r:n};\theta)$ are, when $Y_{r:n}$ is a central order statistic. From (5.1) and (5.2), on approximating the sums by integrals it can be shown that

$$\mathcal{S}(Y_{r:n};\theta) \cong \frac{n(1-p)\{-\log(1-p)\}^2}{\theta^2 p}, \tag{5.5}$$

where $p=r/n$, is away from either 0 or 1. Similarly, from (5.4) one obtains (Arnold *et al.* (1992), p. 166),

$$\mathcal{I}(Y_{r:n};\theta) \cong \frac{2}{\theta^2} + \frac{n(1-p)\{-\log(1-p)\}^2}{\theta^2 p}. \tag{5.6}$$

Thus, $\mathcal{S}(Y_{r:n};\theta)$ and $\mathcal{I}(Y_{r:n};\theta)$ are indeed very close for large n. From (5.5) and (5.6) it is clear that both of them monotonically increase in $(0,p_0)$ and then decrease, where p_0, the unique solution of $2p+\log(1-p)=0$ in $(0,1)$, is approximately 0.7968. Thus, we can conclude that, roughly speaking, for large samples, the TLS (or the FIM) of a single order statistic increases up to the 80th sample percentile and then decreases. This provides a specific illustration of the general statement made at the end of Section 4 (see Remark (3)). The TLS of the 80th percentile is about $0.65n$ times that of a single observation.

For small samples the maximum TLS may not be at the 80th percentile; for example, $\mathcal{S}(Y_{5:5};\theta) > \mathcal{S}(Y_{4:5};\theta)$.

Example 3. (The Normal Distribution) Suppose the population is $N(\mu,\sigma^2)$, so that the location parameter is μ and the scale parameter is σ. Define

$$\nu_1 = \boldsymbol{\alpha}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\alpha}/\nu, \quad \nu_2 = \mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1}/\nu \quad \text{and} \quad \nu_3 = \mathbf{1}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\alpha}/\nu, \tag{5.7}$$

where $\nu = (\boldsymbol{\alpha}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\alpha})(\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1}) - (\mathbf{1}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\alpha})^2$. Let $\hat{\mu}$ and $\hat{\sigma}$ be the BLUE's of μ and σ based on order statistics, respectively. It is known from Lloyd (1952) that

Table 1. TLS and FIM about μ in $\boldsymbol{Y} = (Y_{r:n}, \ldots, Y_{k:n})$ from $N(\mu, 1)$ population ($n = 10$).

r	k	$\mathcal{S}(\boldsymbol{Y};\mu)$	$\mathcal{I}(\boldsymbol{Y};\mu)$	r	k	$\mathcal{S}(\boldsymbol{Y};\mu)$
1	1	2.904	3.110	2	3	6.004
2	2	4.662	4.763	2	4	7.006
3	3	5.714	5.755	2	5	7.786
4	4	6.332	6.345	2	6	8.399
5	5	6.620	6.622	2	7	8.899
1	2	4.856	4.969	2	8	9.282
1	3	6.203	6.260	2	9	9.588
1	4	7.206	7.238	3	4	6.709
1	5	7.992	8.005	3	5	7.486
1	6	8.612	8.617	3	6	8.105
1	7	9.102	9.106	3	7	8.595
1	8	9.492	9.495	3	8	8.985
1	9	9.793		4	5	7.101
1	10	10	10	4	6	7.716
				4	7	8.210
				5	6	7.231

$\mathrm{Var}(\hat{\mu}) = \nu_1\sigma^2$, $\mathrm{Var}(\hat{\sigma}) = \nu_2\sigma^2$, and $\mathrm{Cov}(\hat{\mu}, \hat{\sigma}) = -\nu_3\sigma^2$. For the normal parent, Table 10C.2 of Sarhan and Greenberg (1962) enumerates these moments for all censored samples up to size 20.

With $\theta = \mu$ and $\sigma = 1$, $\mathcal{S}(\boldsymbol{Y};\theta) = \mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1}$ from (4.2), and with $\mu = 0$ and $\theta = \sigma$, $\mathcal{S}(\boldsymbol{Y};\theta) = (\boldsymbol{\alpha}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\alpha})/\theta^2$ from (4.4). One can use the tabulated values of $\boldsymbol{\alpha}$ and $\boldsymbol{\Sigma}$ available in the literature (for example, in Teichroew (1962), Table 10B.1, Table 10B.3) to compute $\mathcal{S}(\boldsymbol{Y};\theta)$ in both the cases. But the computation of $\boldsymbol{\Sigma}^{-1}$ is rather messy here. However, note that $\mathcal{S}(\boldsymbol{Y};\theta) = \nu \cdot \nu_2$ for the location parameter and $\mathcal{S}(\boldsymbol{Y};\theta) = \nu \cdot \nu_1/\theta^2$ for the scale parameter, where the ν_i's are given in (5.7). We have used these facts in the computation of $\mathcal{S}$ in the two tables presented here.

Table 1 exhibits the values of $\mathcal{S}(\boldsymbol{Y};\mu)$ for $n = 10$ and for various choices of r and k, assuming $\sigma = 1$. Nagaraja (1983) has computed $\mathcal{I}(Y_{r:n};\mu)$ and Mehrotra *et al.* (1979) have computed $\mathcal{I}(\boldsymbol{Y};\mu)$ for $r = 1$, $k = 2(1)8$. For the sake of comparison, we also include these values in our table. It shows that $\mathcal{S}(\boldsymbol{Y};\mu)$ is pretty close to $\mathcal{I}(\boldsymbol{Y};\mu)$. Further, they are equal when $\boldsymbol{Y}$ represents the entire sample. This also follows from the fact that $\boldsymbol{Y}$ then belongs to an exponential family. The central order statistics contain more information on μ than the extreme ones. Just the two middle order statistics account for about 72% of the TLS in the whole sample. This is also the TLS of the sample median.

Table 2 exhibits the values of $\sigma^2\mathcal{S}(\boldsymbol{Y};\sigma)$ for order statistics from a random sample of size 10 from $N(0,\sigma^2)$ distribution. Values of $\sigma^2\mathcal{I}(\boldsymbol{Y};\sigma)$ from Mehrotra *et al.* (1979) are also presented for the purpose of comparison. (The blank cells in that column indicate that the corresponding $\mathcal{I}(\boldsymbol{Y};\sigma)$ are unavailable.) As to be

Table 2. TLS and FIM about σ in $\boldsymbol{Y} = (Y_{r:n}, \ldots, Y_{k:n})$ from $N(0,1)$ population ($n = 10$).

r	k	$\mathcal{S}(\boldsymbol{Y};\sigma)$	$\mathcal{I}(\boldsymbol{Y};\sigma)$	r	k	$\mathcal{S}(\boldsymbol{Y};\sigma)$
1	1	6.876		2	3	4.687
1	2	7.305	8.572	2	4	4.961
1	3	7.316	9.047	2	5	5.622
1	4	7.585	9.604	2	6	6.674
1	5	8.245	10.444	2	7	8.120
1	6	9.301	11.622	2	8	9.932
1	7	10.740	13.152	2	9	12.136
1	8	12.550	15.041	3	4	2.744
1	9	14.739		3	5	3.412
1	10	17.361	20	3	6	4.473
2	2	4.674		3	7	5.920
3	3	2.460		3	8	7.740
4	4	0.894		4	5	1.573
5	5	0.100		4	6	2.643
				4	7	4.095
				5	6	1.182

anticipated, extreme order statistics have more information on σ than the central ones. Further, numerically, the TLS of the entire sample is about 87% of the FIM. This implies that the BLUE of σ is 87% efficient.

Acknowledgements

I am indebted to two anonymous referees for their thorough reading of the manuscript and for several helpful comments.

References

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd ed., Wiley, New York.

Arnold, B. C., Balakrishnan, N. and Nagaraja, H. N. (1992). *A First Course in Order Statistics*, Wiley, New York.

Balakrishnan, N. and Cohen, A. C. (1991). *Order Statistics and Inference—Estimation Methods*, Academic Press, Boston.

David, H. A. (1981). *Order Statistics*, 2nd ed., Wiley, New York.

Ferentinos, K. and Papaioannou, T. (1981). New parametric measures of information, *Inform. and Control*, **51**, 193–208.

Lehmann, E. L. (1983). *Theory of Point Estimation*, Wiley, New York.

Lloyd, E. H. (1952). Least-squares estimation of location and scale parameters using order statistics, *Biometrika*, **39**, 88–95.

Mehrotra, K. G., Johnson, R. A. and Bhattacharyya, G. K. (1979). Exact Fisher information for censored samples and the extended hazard rate functions, *Comm. Statist. Theory Methods*, **8**, 1493–1510.

Mosteller, F. (1946). On some useful 'inefficient' statistics, *Ann. Math. Statist.*, **17**, 377–407.

Nagaraja, H. N. (1983). On the information contained in an order statistic, Tech. Report, 278, Dept. of Statistics, The Ohio State University.

Ogawa, J. (1951). Contributions to the theory of systematic statistics, I, *Osaka Mathematical Journal*, **3**, 175–213.

Ogawa, J. (1962). Estimation of the location and scale parameters by sample quantiles (for large samples), *Contributions to Order Statistics* (eds. A. E. Sarhan and B. G. Greenberg), 47–55, Wiley, New York.

Pratt, J. W. and Gibbons, J. D. (1981). *Concepts of Nonparametric Theory*, Springer, New York.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd ed., Wiley, New York.

Sarhan, A. E. and Greenberg, B. G. (1962). The best linear estimates for the parameters of the normal distribution, *Contributions to Order Statistics* (eds. A. E. Sarhan and B. G. Greenberg), 206–269, Wiley, New York.

Takahashi, H. and Sugiura, N. (1989). The rate of convergence of Fisher information for type II censored sample, *J. Japan Statist. Soc.*, **19**(2), 139–144.

Teichroew, D. (1962). Tables of lower moments of order statistics for samples from the normal distribution, *Contributions to Order Statistics* (eds. A. E. Sarhan and B. G. Greenberg), 190–205, Wiley, New York.

Tukey, J. W. (1965). Which part of the sample contains the information?, *Proc. Nat. Acad. Sci. U.S.A.*, **53**, 127–134.

Wijsman, R. A. (1973). On the attainment of the Cramér-Rao lower bound, *Ann. Statist.*, **1**, 538–542.